



Technical note: Reconstructing missing surface aerosol elemental carbon data in long-term series with ensemble learning

Qingxiao Meng¹, Yunjiang Zhang¹, Sheng Zhong², Jie Fang¹, Lili Tang¹, Yongcai Rao³, Minfeng Zhou⁴, Jian Qiu⁵, Xiaofeng Xu⁶, Jean-Eudes Petit⁷, Olivier Favez⁸, and Xinlei Ge¹

¹Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

²Jiangsu Environmental Monitoring Center, Nanjing, Nanjing 210019, China

³Xuzhou Environmental Monitoring Center of Jiangsu, Xuzhou 221018, China

⁴Suzhou Environmental Monitoring Center of Jiangsu, Suzhou 215000, China

⁵Zhenjiang Environmental Monitoring Center of Jiangsu, Zhenjiang 212000, China

⁶Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044, China

⁷Laboratoire des Sciences du Climat et de l'Environnement, CEA/Orme des Merisiers, Gif-sur-Yvette, 91191, France

⁸Institut National de l'Environnement Industriel et des Risques, Verneuil-en-Halatte, 60550, France

Correspondence: Yunjiang Zhang (yjzhang@nuist.edu.cn)

Received: 4 September 2024 – Discussion started: 25 November 2024

Revised: 6 April 2025 – Accepted: 10 April 2025 – Published: 15 July 2025

Abstract. Ground-based measurements of elemental carbon (EC) – classified under thermal–optical methods and considered a surrogate for black carbon – are essential for assessing air quality and evaluating climate impacts. However, data gaps caused by technical challenges impede comprehensive analyses of long-term trends. This study proposed an ensemble learning modeling method to address these challenges. The model used readily accessible ground observation air pollutant data as proxies for EC-related tracers, along with meteorological parameters, to enhance prediction accuracy. It integrated outputs from Gradient Boosting Regression Trees, eXtreme Gradient Boosting, and random forest models, combining them through ridge regression to produce robust predictions. We applied this approach to reconstruct hourly EC concentrations from 2013–2023 for four cities in eastern China, filling 45 %–79 % of missing data and improving prediction performance by 8 %–17 % compared to individual models. Over the 11-year period, EC exhibited an overall decline (−0.20 to −0.14 μg m^{−3} a^{−1}), with a more significant decrease from 2013–2020 (−0.24 to −0.15 μg m^{−3} a^{−1}). During this time, the average EC concentration in the four cities dropped from 3.26 to 1.59 μg m^{−3}, followed by a noticeable slowdown in the rate of decline from 2020–2023 (−0.12 to −0.04 μg m^{−3} a^{−1}). Additionally, a fixed emission approximation method based on ensemble learning was proposed to quantitatively analyze the drivers of long-term EC trends. The analysis revealed that anthropogenic emission controls were the predominant contributors, accounting for approximately 92 % of the changes in EC trends from 2013–2020. However, their influence weakened post-2020, contributing approximately 80 %. These findings highlighted that while China's Clean Air Actions implemented since 2013 have substantially reduced black carbon concentrations, sustained and enhanced strategies are still necessary to further mitigate black carbon pollution.

1 Introduction

Black carbon (BC), derived from the incomplete combustion of fossil or biomass fuels (Bond and Bergstrom, 2006), is a significant component of fine particulate matter ($\text{PM}_{2.5}$) in the troposphere. BC has a strong capacity to absorb visible light (Bond et al., 2013), which can directly or indirectly affect global climate change (Ramanathan and Carmichael, 2008) and influence the structure of the urban boundary layer, exacerbating regional air pollution formation (Ding et al., 2016). Additionally, BC particles, which are small in size and rich in toxic substances on their surfaces, pose significant health risks (Valavanidis et al., 2013), leading to respiratory and cardiovascular diseases (Wei et al., 2023). Therefore, establishing long-term trends of BC is crucial for studying its climate, environmental, and health impacts.

Methods for determining ground-level BC concentrations and long-term trends generally include ground-based in situ observations (Peng et al., 2019; Wei et al., 2020), satellite retrievals (Yu et al., 2024; Zhao et al., 2021), and atmospheric chemical transport model simulations (Yang et al., 2021). Satellite-based retrieval methods are often used to characterize ground-level BC concentrations. For instance, Li et al. (2022) used multi-angle polarization satellite observations to retrieve the spatiotemporal distribution of global BC concentrations. However, satellite data are susceptible to noise originating from clouds and other factors, leading to data gaps and large uncertainties (Bao et al., 2019; Zhao et al., 2021). For example, Gogoi et al. (2023) compared in situ BC observations in India from 2019–2020 with BC concentrations retrieved by the Cloud and Aerosol Imager-2 (CAI-2) on board the Greenhouse gases Observing Satellite-2 (GOSAT-2) and found a 33 % bias. Beyond measurement uncertainties, satellite retrieval techniques also generally cannot provide long-term, continuous high-time-resolution (e.g., hourly) datasets.

To address the limitations of satellite retrievals, reanalysis datasets have developed, combining various in situ, satellite observations and short-term numerical weather prediction products through data reanalysis techniques. Notable datasets include the ECMWF Re-Analysis Interim (ERA-Interim) (Dee et al., 2011), MODIS Atmosphere and Land Reanalysis (MODIS) (Levy et al., 2013), and Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) (Randles et al., 2017). MERRA-2, in particular, offers convenient high-time-resolution datasets for atmospheric pollutants, including BC (Bali et al., 2017). However, there are uncertainties when comparing ground-based observations to reanalysis data. For instance, several studies found that MERRA-2 overestimated BC concentrations at most Chinese stations by approximately 30 % (Ma et al., 2021; Xu et al., 2020; Yu et al., 2024). Atmospheric chemical transport models are also commonly used to simulate the spatiotemporal distribution of BC concentrations. For example, Matsui (2020) used the Community Atmosphere Model Version 5 (CAM5) and the Aerosol Two-dimensional bin mod-

ule for foRmation and Aging Simulation (ATRAS) model to simulate the increase in global BC concentrations from pre-industrial times to the present. However, uncertainties in parameterization schemes for some physical and chemical processes in numerical models remain (Ervens, 2015; Harrison, 2018). Additionally, emission inventories used in models have inherent uncertainties and may not be updated promptly, affecting simulation results (Xu et al., 2021).

In situ observations remain the most direct and effective method for quantifying BC concentrations. In practice, elemental carbon (EC), measured using thermal–optical methods, is often employed as a surrogate for BC (Bond et al., 2013), especially in cases where optical methods are unavailable. These ground-based measurements (EC and/or BC) are central to long-term observational networks worldwide, such as the Atmospheric Science and Chemistry Measurement Network (ASCENT) in the United States (Ng et al., 2022); the Aerosol, Clouds, and Trace Gases Research Infrastructure (ACTRIS) in Europe (Laj et al., 2024); and the China Atmosphere Watch Network (CAWNET) in China (Zhang et al., 2014). While these methods enable the establishment of time series for BC or EC concentrations, they lack the capacity to provide historical data without prior measurements. Furthermore, practical challenges such as instrument malfunctions, routine maintenance, and hardware limitations at observation sites often lead to data gaps. These interruptions hinder the continuity and completeness of long-term datasets, posing significant challenges for trend analysis and comprehensive assessments.

This study introduces an ensemble learning method leveraging ground-based observational data (including in situ EC and air pollutant measurements), BC column concentration assimilation data, and meteorological datasets. Applying this approach, we successfully reconstructed hourly EC concentration time series for four representative cities in the Yangtze River Delta region in eastern China from 2013–2023. Furthermore, an ensemble learning model was developed to evaluate the drivers of EC trends, enabling a quantitative analysis of the relative contributions of anthropogenic emission reductions and meteorological variations to the EC trends in these cities over the 11-year period.

2 Data and methods

2.1 Air pollutant and meteorological data

The observation sites for this study are located in the representative cities of Nanjing, Suzhou, Xuzhou, and Zhenjiang within the Yangtze River Delta city cluster in eastern China. All these sites are urban monitoring stations, and the coordinates of the sampling sites are provided in Table S1 in the Supplement. Gaseous pollutants at the corresponding sites in the four cities, including carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2), were sourced from the corresponding sampling sites. The meteorological data

used in this study are listed in Table S2 in the Supplement. These meteorological conditions are derived from the ERA5 reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF). The spatial and temporal resolutions of ERA5 data are $0.25^\circ \times 0.25^\circ$ and 1 h, respectively. To represent the meteorological conditions at the observation sites, we extracted data from the ECMWF Reanalysis v5 (ERA5) grid cells that correspond to the coordinates of the monitoring stations.

2.2 Measurements and inter-comparison of BC and EC

In this study, long-term measurements of EC were conducted across four cities (i.e., Nanjing, Suzhou, Xuzhou, and Zhenjiang), using the Sunset Laboratory semi-continuous OC/EC analyzer (Model-4), which provided hourly time resolution. The analysis employed the National Institute for Occupational Safety and Health (NIOSH) thermal–optical transmittance (TOT) method to quantify EC. The thermal–optical approach, which forms the basis of the OC/EC analyzer, is detailed extensively in previous studies (Arhami et al., 2006; Birch and Cary, 1996; Jung et al., 2011; Zheng et al., 2014). In addition to EC measurements, refractory black carbon (rBC) data from our previous study in Nanjing (Yang et al., 2019) were used for inter-comparison with EC data. The rBC measurements were obtained using a single-particle soot photometer (SP2) (Liu et al., 2010; Cross et al., 2010). Briefly, the SP2 operates on the principle of laser-induced incandescence (Liu et al., 2010), which involves heating individual rBC particles to high temperatures using a focused laser beam. The incandescence signal emitted during this process is then used to quantify rBC concentration, based on the characteristic heating curve of rBC particles (Liu et al., 2010). Figure S1 in the Supplement shows the relationship between rBC and EC mass concentrations for the Nanjing dataset in this study, along with those from previous work (Pileci et al., 2021). The results revealed a good agreement between rBC and EC (slope = 1.01) in Nanjing, which is consistent with the range reported by Pileci et al. (2021).

2.3 BC data from reanalysis or simulation dataset

The black carbon column mass density (BCC) and black carbon surface mass concentration (BCS) data used in this study were obtained from MERRA-2 (M2T1NXADG, V5.12.4). MERRA-2 is a new reanalysis dataset that was released by the NASA Global Modeling and Assimilation Office (GMAO) in 2017. The spatial resolution of the data used in this study is $0.5^\circ \times 0.625^\circ$, and the temporal resolution is 1 h. The BC concentrations in MERRA-2 are estimated by assimilating satellite-derived aerosol optical depth (AOD) into an atmospheric chemical transport model (Gelaro et al., 2017). Since satellite retrievals are based on optical properties, the inferred BC concentrations can be affected by optical effects, such as the “lensing effect” (Liu et al., 2015), which

could potentially lead to an overestimation of BC concentrations. To compare this with the BC concentration simulated by the atmospheric chemical transport model approach, we used the Tracking Air Pollution in China (TAP) dataset (<http://tapdata.org.cn/>, last access: 20 August 2024) (Geng et al., 2021). In brief, the TAP dataset includes the surface BC concentration data, which is simulated by the Community Multiscale Air Quality (CMAQ) model and machine learning method (Geng et al., 2021; Liu et al., 2022). For simplicity, TAP BC did not account for methodological differences between BC and EC measurements (Liu et al., 2022). As a result, these BC data could be influenced by factors such as model assumptions, emission inventories, and meteorological conditions (Yu et al., 2024; Liu et al., 2022).

2.4 Ensemble learning model

In this study, we applied the ensemble learning model (EL) approach to address the two major issues. First, we developed a model to reconstruct long-term trend of EC at urban observation sites, filling in missing data. Second, we proposed an ensemble learning approach to evaluate the driving factors of EC trends, which can quantify the contributions of emission reduction and meteorological variation to the EC trends. The two modeling methodologies integrate the predictions of Gradient Boosting Regression Trees (GBRTs), eXtreme Gradient Boosting (XGBoost), and random forest (RF) using ridge regression. GBRTs and XGBoost iteratively train decision tree models, respectively, which reduces residuals step by step to make predictions. The final prediction results are weighted sums of the predictions from each tree model, with different weights for each tree. The RF model consists of multiple decision trees, each of which provide a prediction. The prediction method averages the results from each decision tree to obtain the prediction output, with each tree having equal weight.

For multivariate regression analysis, we chose ridge regression over traditional multiple linear regression to account for multicollinearity among the three model outputs. Ridge regression is particularly effective in handling multicollinearity by introducing a regularization term that improves computational stability and reduces the risk of overfitting (Kidwell and Brown, 1982; Hoerl and Kennard, 1970). The final ensemble learning model, integrating the results through ridge regression to determine coefficients (m_1 , m_2 , and m_3) for each machine learning model, is given by Eq. (1):

$$f_{\text{EL}} = m_1 f_{\text{GBRTs}} + m_2 f_{\text{XGBoost}} + m_3 f_{\text{RF}}. \quad (1)$$

2.4.1 Reconstructing missing data of EC

Reconstructing long-term EC data involved integrating hourly meteorological variables and emission indicators into an ensemble learning model. Individual ensemble learning models were established for each city. As shown in Table S2,

the meteorological variables include 18 factors. The emission indicators include BCC and in situ surface observations of CO, SO₂, and NO₂. These air pollutants, along with EC, are mainly associated with fuel combustion processes: CO concentration is closely linked to combustion source activities (such as agricultural crops, forest fires, and fossil fuel) (Reid et al., 2005; Wang et al., 2011). SO₂ is generally associated with industrial activities (such as coal combustion) (Wang and Chen, 2016), while NO₂ primarily originates from vehicle emissions (Carslaw, 2005). All of them could thereby be indirectly indicating EC sources.

In this modeling approach, available in situ EC observation data for each city, along with corresponding meteorological and emission indicator variables, were used to construct the training and testing datasets. The training set was used to develop the ensemble learning model for data reconstruction, while the testing set served to validate model performance and determine optimal model parameters. Once the model was optimized, all input parameters from 2013–2023 were fed into the ensemble learning model to reconstruct long-term EC data. As shown in Fig. S2 in the Supplement, EC observations in Nanjing included approximately 49 000 valid data points, while approximately 90 000 points were reconstructed, resulting in a data completion rate of approximately 47 %. The completion rates for the other cities were approximately 45 % for Suzhou, 50 % for Xuzhou, and 79 % for Zhenjiang, respectively.

2.4.2 Driver analysis of long-term trends

The analysis of drivers behind EC trends differs from the reconstruction of missing EC data by excluding emission indicator variables from the prediction features. To quantify the factors influencing long-term reconstructed EC trends, we developed a machine-learning based fixed emission approximation (FEA) method. Using the reconstructed long-term EC data, we assume each year i ($i = 2013, 2014, \dots, 2023$) could serve as a baseline year for the initial anthropogenic emission conditions, with data from the chosen year used to train the model. Reconstructed EC data and meteorological variables from the selected baseline year are employed as training and testing datasets to build ensemble learning models specific to individual cities. These models are subsequently applied to predict EC concentrations for the entire period from 2013–2023. The resulting predictions reflect variations driven solely by meteorological conditions, assuming fixed emissions at baseline-year levels. This approach effectively isolates the contributions of meteorological variations from changes driven by emissions.

The difference in the observed EC concentrations ($\Delta\text{OBS}_{(j,k)}$) between 2 different years (j, k) is jointly influenced by the inter-annual relative changes in EC concentrations driven by meteorological variables ($\Delta\text{MET}_{(j,k)}$) and anthropogenic emissions ($\Delta\text{ANT}_{(j,k)}$). This relationship is

described mathematically in Eq. (2) as follows:

$$\Delta\text{OBS}_{(j,k)} = \Delta\text{ANT}_{(j,k)} + \Delta\text{MET}_{(j,k)}. \quad (2)$$

This equation (Eq. 2) assumes that changes in meteorology and emissions are the primary factors affecting the observed EC concentration variations between years, allowing for a decomposition of their respective contributions to the trends. The $\Delta\text{MET}_{(j,k)}$ can be estimated by comparing the predicted EC concentrations for 2 different years ($C_{\text{MET}(i,j)}$ and $C_{\text{MET}(i,k)}$), where j and k represent the years for which predictions are made using the model trained with data from baseline year i . Specifically, $j = 2013, 2014, \dots, 2022$, and $k = 2014, 2015, \dots, 2023$, and $k > j$. This relationship is expressed in Eq. (3), which isolates the influence of meteorological variations on inter-annual differences in EC concentrations, assuming emissions at the baseline year i level.

$$\Delta\text{MET}_{(j,k)} = C_{\text{MET}(i,k)} - C_{\text{MET}(i,j)}. \quad (3)$$

Thus, the impact of anthropogenic emission controls on changes in EC concentrations can be determined by subtracting $\Delta\text{MET}_{(j,k)}$, which reflects the contribution of meteorological variations, from $\Delta\text{OBS}_{(j,k)}$, the observed inter-annual EC concentration change. This relationship is expressed in Eq. (4):

$$\Delta\text{ANT}_{(j,k)} = \Delta\text{OBS}_{(j,k)} - \Delta\text{MET}_{(j,k)}. \quad (4)$$

Here, $\Delta\text{ANT}_{(j,k)}$ represents the portion of the EC concentration change attributable to anthropogenic emission controls during the period between years j and k .

Based on the FEA method, we further proposed an analytical approach to assess the impact of short-term pandemic lockdowns. Specifically, we analyzed the relative contribution of the COVID-19 lockdown (LD) period, which lasted from 26 January–17 February 2020 (Huang et al., 2021). The calculation of the impact of the 2020 LD on pollutant concentrations is determined by the proportion of the LD period samples within the entire year of 2020 (r_{LD}), as well as the difference between observed values (OBS_{LD}) and predicted values ($C_{\text{MET}(i,\text{LD})}$) during the pandemic period. This allows us to calculate the total contribution of the COVID-19 lockdown and the emission control effects between 2020 and the baseline modeling year on pollutants. Assuming that the contribution of anthropogenic emissions to pollutants during the winter of 2019 and other years remains stable, this can be calculated by the difference between observed values (OBS_{NLD}) and predicted values ($C_{\text{MET}(i,\text{NLD})}$) during the remaining time in 2020. Therefore, the contribution of the 2020 LD to pollutants is represented by Eq. (5):

$$\text{COVID-19}_{(i,\text{LD})} = r_{\text{LD}}(\text{OBS}_{\text{LD}} - C_{\text{MET}(i,\text{LD})}) - r_{\text{LD}}(\text{OBS}_{\text{NLD}} - C_{\text{MET}(i,\text{NLD})}). \quad (5)$$

2.4.3 Model evaluation and uncertainty analysis

In this study, the model performance was assessed by comparing the reconstructed results for the test set with observa-

tional data. To evaluate the deviation between the observed ground-based values (y_{obs}) and the reconstructed predicted values (y_{pre}), we used several statistical metrics: root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), and the correlation coefficient (R), as defined by the following equations.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{obs}} - y_{\text{pre}})^2}, \quad (6)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_{\text{obs}} - y_{\text{pre}})^2, \quad (7)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{\text{obs}} - y_{\text{pre}}|. \quad (8)$$

To validate and assess the model's performance in reconstructing missing data, we evaluated the prediction error metrics on the test set, including RMSE, MSE, MAE, and R . Figures 1 and Fig. S3 in the Supplement summarize the performance metrics for individual models, including XGBoost ($R = 0.81 \pm 0.05$), GBRTs ($R = 0.84 \pm 0.03$), and RF ($R = 0.87 \pm 0.03$), as well as for the ensemble learning model. The ensemble machine learning model exhibited a notable improvement, achieving $R = 0.94 \pm 0.01$, which represents a 17 % enhancement over the XGBoost model. Additionally, MAE, MSE, and RMSE decreased by 47 %, 67 %, and 42 %, respectively. These results highlight the superior performance of the ensemble machine learning model, showcasing its robustness and accuracy in reconstructing missing data. As shown in Table S3 in the Supplement, the ensemble model also yielded the best performance within the FEA framework. We further evaluated the importance of the MERRA-2 black carbon concentration (BCC) as a predictor by testing the model's performance both with and without this variable (see Fig. S4 in the Supplement). Inclusion of MERRA-2 BCC significantly improved the model's performance across all evaluation metrics, confirming it as a key contributor to model accuracy.

To further assess the influence of emission changes and test model robustness, we trained an alternative version of the model excluding all emission indicator variables. Compared to the full model, this version showed a marked degradation in performance (see Fig. S5 in the Supplement). This result highlights the critical role of including emission-related variables in ensuring accurate and reliable predictions. Additionally, as illustrated in Fig. S6 in the Supplement, the CO/NO₂ ratio during periods with valid EC observations closely resembled that during periods with missing EC data. This consistency suggests that interannual variation in these emission indicator ratios likely exerted minimal influence on the model's predictive performance. Finally, long-term trends in monthly mean EC or BC concentrations were assessed using the non-parametric Mann–Kendall (MK) trend test. To account for seasonal variability, trend slopes were derived using the seasonal Theil–Sen estimator, enhancing robustness in the presence of periodic fluctuations.

To evaluate the relative difference in the analysis of EC trend change drivers, we proposed a method for quantifying the relative difference in the FEA. The term $\Delta\text{ANT}_{(i,j)}$ is determined by subtracting $\Delta\text{MET}_{(i,j)}$ from $\Delta\text{OBS}_{(i,j)}$. By substituting Eq. (3) from this paper into this calculation, the relationship is expressed as shown in Eq. (9). The term $C_{\text{MET}(i,i)}$ was the self-prediction for the year i based on the training year i . This formulation allows for a systematic evaluation of the uncertainties associated with the FEA approach, ensuring robust attribution of trends to anthropogenic emission controls and meteorological variations. Similarly, by substituting results for the 2 years i and j , we can obtain $\Delta\text{ANT}_{(j,i)}$, as shown in Eq. (10).

$$\Delta\text{ANT}_{(i,j)} = \Delta\text{OBS}_{(i,j)} - (C_{\text{MET}(i,j)} - C_{\text{MET}(i,i)}) \quad (9)$$

$$\Delta\text{ANT}_{(j,i)} = \Delta\text{OBS}_{(j,i)} - (C_{\text{MET}(j,i)} - C_{\text{MET}(j,j)}) \quad (10)$$

If the FEA method were entirely free of relative difference, the relationship $(C_{\text{MET}(i,j)} - C_{\text{MET}(i,i)}) + (C_{\text{MET}(j,i)} - C_{\text{MET}(j,j)}) = 0$ would hold true, implying that $\Delta\text{ANT}_{(i,j)} + \Delta\text{ANT}_{(j,i)} = 0$. However, as with any method, some degree of relative difference could be unavoidable. To account for this, the relative difference of the data for year j , predicted using year i as the training data ($Y_{i(j)}$), was determined by the absolute value of the sum of $\Delta\text{ANT}_{(i,j)}$ and $\Delta\text{ANT}_{(j,i)}$, and then normalizing by $C_{\text{MET}(i,j)}$. This calculation is expressed in Eq. (11):

$$Y_{i(j)} = \frac{|\Delta\text{ANT}_{(i,j)} + \Delta\text{ANT}_{(j,i)}|}{C_{\text{MET}(i,j)}}. \quad (11)$$

This approach provided a quantifiable measure of relative difference inherent in the FEA method, facilitating a more robust evaluation of the predictions. When $i = j$, $\Delta\text{ANT}_{(i,j)} = \Delta\text{ANT}_{(j,i)} = 0$, and the relative difference calculation at this point is expressed by Eq. (12).

$$Y_{i(i)} = \frac{|\text{OBS}_i - C_{\text{MET}(i,i)}|}{C_{\text{MET}(i,i)}} \quad (12)$$

As shown in Fig. 2, the average relative difference across the four cities over the 11-year period is approximately $4 \pm 4\%$. The lowest relative difference was observed in Xuzhou, with $3 \pm 3\%$, while the highest was in Suzhou, with $6 \pm 5\%$. Because this study incorporates reconstructed EC data, potential discrepancies between model-derived training data and actual observations may introduce additional uncertainty into the results. To assess this, we selected Nanjing as a representative test case. Specifically, we reconstructed the 2013–2023 EC dataset using training data from two distinct time periods: 2013–2020 and 2014–2019. We then applied the FEA method to both reconstructions and compared the outcomes. As shown in Fig. S7 in the Supplement, while some variations exist between the two sets of results, the average difference remains within approximately 10 %. This finding suggests that the choice of training dataset can introduce a

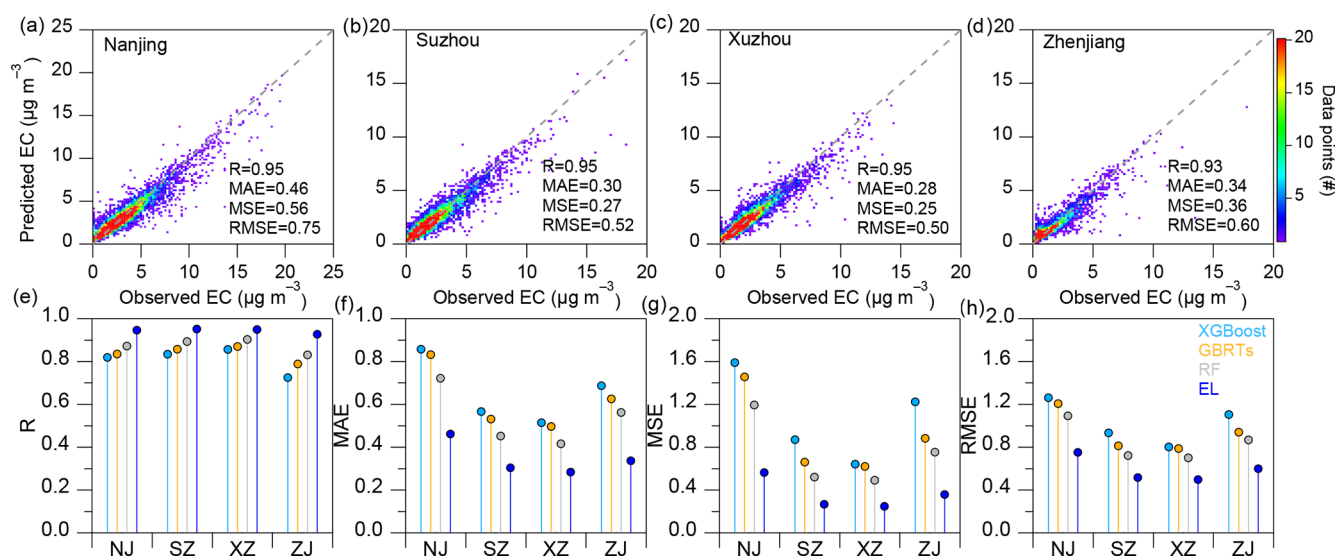


Figure 1. Performance evaluation of the models for reconstructing hourly EC concentration. **(a–d)** Comparison of the ensemble learning model-predicted concentration and observed EC concentration. The legend indicates the number of data points at each binning interval, which is approximately $0.25 \mu\text{g m}^{-3}$. **(e–h)** Comparison of the model performance parameters for four cities, i.e., Nanjing (NJ), Suzhou (SZ), Xuzhou (XZ), and Zhenjiang (ZJ).

moderate degree of uncertainty to the FEA results – an inherent characteristic of ensemble learning and other statistical modeling approaches. Nonetheless, the relatively small magnitude of this difference reinforces the reliability and generalizability of the machine-learning-based FEA framework. In Sect. 3.2, we further compare the FEA method with the widely used de-weathered approach proposed by Grange et al. (2018), to evaluate the consistency and applicability of different trend attribution frameworks.

3 Results and discussion

3.1 Reconstruction of missing data of EC and comparison

As illustrated in Fig. 3a–d, the reconstructed EC concentrations in all four cities closely track the trends observed in the ground-based measurements. These reconstructed values also exhibit strong agreement with the temporal patterns of TAP-derived BC concentrations. Over the 11-year study period, both reconstructed EC and TAP BC showed statistically significant and consistent decreasing trends, with rates ranging from -0.20 to $-0.14 \mu\text{g m}^{-3} \text{a}^{-1}$ and -0.22 to $-0.14 \mu\text{g m}^{-3} \text{a}^{-1}$, respectively ($P < 0.05$). In contrast, the MERRA-2 BC product displays weaker or non-significant declining trends, particularly in Suzhou and Xuzhou, where the trend was not statistically significant ($P > 0.05$). To investigate this inconsistency, we applied the FEA method to isolate the meteorological contributions to the observed trends in MERRA-2 BC. As shown in Fig. S8 in the Supplement, MERRA-2 BC trends exhibit good agreement

with the meteorological-driven BC trend derived from the FEA method, suggesting that the interannual variability in MERRA-2 BC was largely driven by meteorological factors. In contrast, the TAP BC dataset showed clear downward trends that were more closely aligned with the changes observed in our reconstructed EC data (Fig. 3b and c). This indicated that the TAP dataset for the two cities was more sensitive to the anthropogenic emission reductions and better captured their influence on BC concentration trends in recent years. Additionally, MERRA-2 BC consistently overestimates EC concentrations compared to both ground-based measurements and reconstructed data. This overestimation has also been reported in previous studies (e.g., Xu et al., 2020), which documented substantial overpredictions of MERRA-2 BC in urban areas in the Yangtze River Delta of eastern China, such as Shanghai and Hangzhou.

Figure 3e–h further demonstrate that the reconstructed EC data capture monthly variations with high fidelity relative to in situ observations. All four datasets exhibit a pronounced seasonal cycle, characterized by elevated concentrations in autumn and winter and lower levels in spring and summer. This seasonal pattern highlights the need for more stringent BC emission controls during the cold months. Among the datasets, TAP-derived BC closely tracks the monthly variability observed in ground-based measurements, whereas the MERRA-2 BC significantly overestimates concentrations – by approximately 147 % on average. The discrepancies in MERRA-2 estimates likely arise from two key factors. First, the relatively coarse spatial resolution of MERRA-2 ($0.5^\circ \times 0.625^\circ$; Randles et al., 2017) means that a grid cell may span diverse land-use types – including urban, subur-

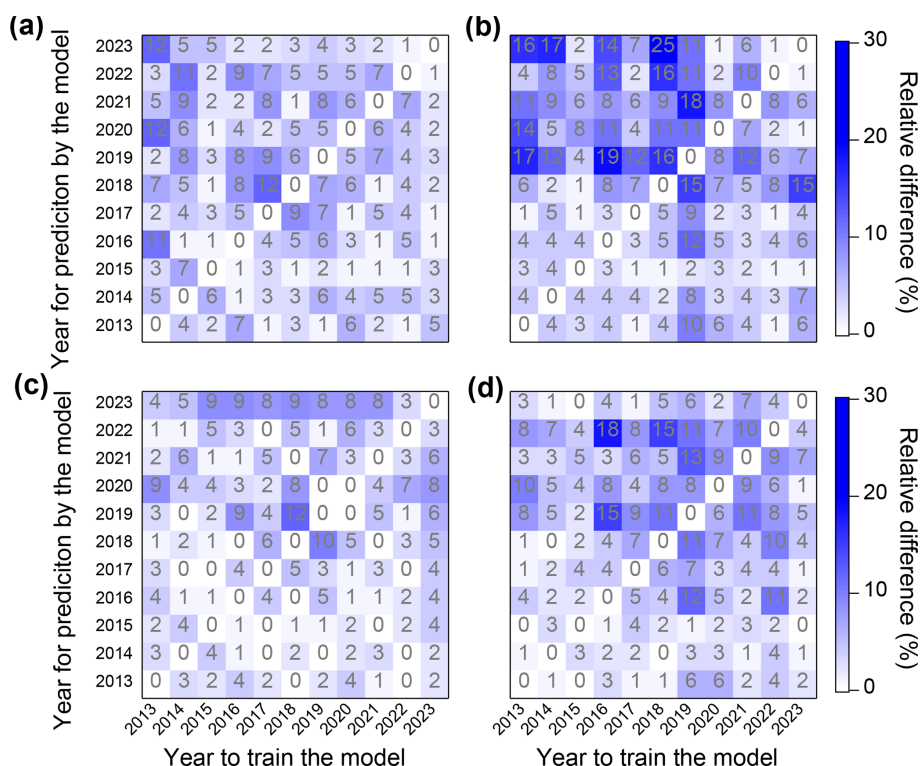


Figure 2. Cross-matrix relative difference analysis of the FEA method: (a) Nanjing, (b) Suzhou, (c) Xuzhou, and (d) Yangzhou. The relative difference here refers to the relative difference of the results obtained from different year to train the model and different year for prediction by the model.

ban, and background regions – thereby smoothing out localized pollution signals and/or inflating background contributions. Second, uncertainties associated with satellite data assimilation – particularly under conditions of severe pollution, cloud cover, or precipitation – may introduce biases in MERRA-2 estimates (Xu et al., 2020), thereby limiting their accuracy in reproducing station-scale EC variability. As shown in Fig. 3i–l and detailed in Table S4 in the Supplement, MERRA-2 BC still maintains a moderate correlation with in situ observations ($R = 0.65 \pm 0.05$), despite its substantial overestimation. In contrast, the TAP BC exhibited better agreement with ground-based data ($R = 0.69 \pm 0.04$). The highest correspondence, however, was observed for the reconstructed dataset, which achieved an exceptional correlation coefficient of $R = 0.97$. This result could highlight the high fidelity and robustness of the reconstruction approach during periods with available observations.

To assess the quality of the reconstructed missing data, we analyzed its correlation with co-located air pollutants (CO and NO₂). As shown in Fig. S9 in the Supplement, the observed EC concentrations exhibit strong correlations with CO ($R = 0.66 \pm 0.10$) and NO₂ ($R = 0.71 \pm 0.06$) over the entire study period. Similarly, the reconstructed EC concentrations for the missing data demonstrate even better correlations, with R values of 0.80 ± 0.06 for CO and 0.85 ± 0.04

for NO₂ (Fig. S10a–h in the Supplement), respectively. Furthermore, the correlation between the TAP BC and the observed EC concentrations ($R = 0.65 \pm 0.05$) was stronger than that observed for MERRA-2. As shown in Fig. S10m–p, the reconstructed EC concentrations for the missing data also exhibit a good correlation ($R = 0.72 \pm 0.04$) with the TAP BC, with an approximately 11 % difference between the two datasets. The slope between TAP and observed EC was 0.60 ± 0.20 , while the slope between TAP and the reconstructed EC was 0.68 ± 0.06 (see Table S4). These comparisons suggest that the reconstructed data were reasonably accurate when compared to ground-based observations.

Using the reconstructed EC data, we analyzed the trends in EC concentrations across the four cities from 2013–2023 (see Table S5 in the Supplement). The EC concentrations have significantly decreased, with reductions of 61 % in Nanjing and Suzhou, 59 % in Xuzhou, and 47 % in Zhenjiang compared to 2013 levels. Specifically, in Nanjing and Xuzhou, the annual average EC concentrations decreased by 2.60 and $1.88 \mu\text{g m}^{-3}$, respectively, over the 11-year period. This trend aligns with findings by Zhou et al. (2024), who reported similar reductions from 2013–2020 based on observational data. These reductions highlighted the effectiveness of China’s emission reduction policies in mitigating BC in the Yangtze River Delta. In Suzhou and Zhenjiang, EC concentrations

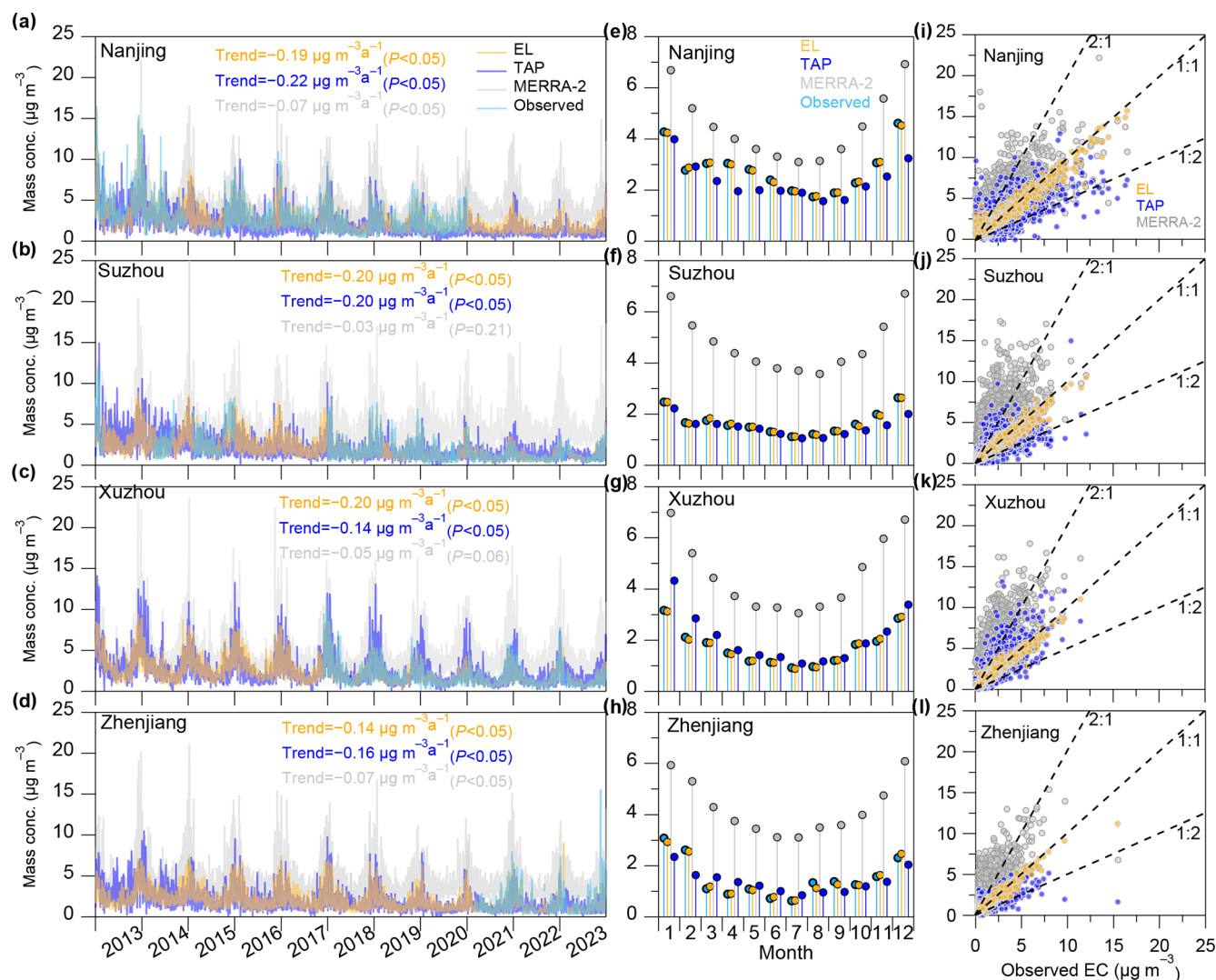


Figure 3. Comparison of EC or BC concentration from different datasets. (a–d) Trends in EC or BC concentration from the four datasets (EL, MERRA-2, TAP, and observation). (e–h) Monthly variations. (i–l) Relationship of reconstructed, MERRA-2, and TAP-modeled EC or BC with observed EC.

decreased by 1.84 and $1.17 \mu\text{g m}^{-3}$, respectively. The average EC concentrations over the 11-year period were $2.01 \pm 0.77 \mu\text{g m}^{-3}$ in Suzhou and $1.98 \pm 0.50 \mu\text{g m}^{-3}$ in Zhenjiang, which were significantly lower than those in the more industrialized cities of Nanjing and Xuzhou.

As shown in Fig. S11a–d in the Supplement, the diurnal variations in EC concentrations in the four cities have significantly decreased over the 11-year period, with peaks during the morning and evening rush hour, primarily due to vehicle emissions. Notably, the rate of EC reduction from 2020–2023 (-0.12 to $-0.04 \mu\text{g m}^{-3} \text{a}^{-1}$) was significantly lower than the overall rate from 2013–2023 (-0.20 to $-0.14 \mu\text{g m}^{-3} \text{a}^{-1}$). Zhao et al. (2024) observed a slowdown in the decline of CO/CO_2 concentrations in the Yangtze River Delta, attributing this trend to limited improvements in the combustion efficiency of anthropogenic sources in re-

cent years. Figures 3a–d and S11a–d further reveal that EC concentrations in the four cities were substantially lower during the COVID-19 pandemic (2020–2022). This aligns with findings by Cui et al. (2021), who reported significant reductions in BC concentrations during the pandemic due to lockdown measures. However, EC concentrations in Nanjing and Suzhou rebounded in 2023, likely due to a resurgence of anthropogenic emissions. Similarly, Liu et al. (2024) observed that global carbon emissions, which temporarily decreased during the pandemic in 2021, rebounded and exceeded previous levels in 2022 and 2023. Despite the significant reductions in EC concentrations across the Yangtze River Delta over the past 11 years, the rate of decrease has slowed in recent years.

3.2 Drivers of the EC trends

To disentangle the effects of meteorology and anthropogenic emissions on EC trends, we also applied a machine-learning-based meteorological normalization method, also known as the de-weathered method (Grange et al., 2018). This method has been widely used to assess the drivers of trends in air pollutants and aerosol chemical composition (Li et al., 2023; Vu et al., 2019; Zhang et al., 2019b; Zhou et al., 2022). The normalization approach developed by Grange et al. (2018) adjusts pollutant concentrations by removing the influence of meteorological variability, thereby isolating the effects of emission control measures. Specifically, the method builds a statistical model to quantify the relationship between pollutant concentrations and meteorological and temporal variables. It then performs 1000 resamplings of historical meteorological data while holding temporal variables constant at their observed values. The model calculates pollutant concentrations across all resampled meteorological scenarios, and the ensemble average of these predictions represents the meteorologically normalized (i.e., de-weathered) concentration. This process enables a more accurate attribution of observed trends to changes in emissions. Detailed methodology can be found in Grange et al. (2018) and related studies (Vu et al., 2019; Zhang et al., 2019b; Zhou et al., 2022). This method can effectively separate the impacts of emissions and meteorology on these trends. Detailed methodologies of this meteorological normalization can be found in previous studies (e.g., Grange et al., 2018; Vu et al., 2019; Zhang et al., 2019b; Zhou et al., 2022). To further validate our proposed FEA method, we compared its results against those from the traditional meteorological normalization approach. As shown in Fig. 4, both methods yielded highly consistent anthropogenic emission-driven trends in EC across the four cities. Additionally, Table S6 in the Supplement shows that all emission-driven trends derived from both methods passed the Mann–Kendall test ($P < 0.05$), with slope differences of less than approximately 8 %. The discrepancies between the two methods remained within approximately 10 %, demonstrating strong agreement and supporting the robustness of the emission-driven trends derived from our FEA method.

To quantify the driving factors behind the observed trends in elemental carbon (EC) – specifically anthropogenic emissions and meteorological variations – we conducted a comprehensive attribution analysis. As shown in Fig. S12 in the Supplement, EC concentrations exhibited a significant decreasing trend from 2013–2023. Meteorological factors played only a minor role, contributing 9 ± 1 % to the observed decline, while anthropogenic emission control measures accounted for the remaining 91 ± 1 %. As illustrated in Fig. S13 in the Supplement, we estimated the meteorology-driven component of EC trends using both the conventional de-weathering method and our proposed FEA method. The results from both approaches indicated that meteorology-driven EC trends remained relatively flat over the 11-year pe-

riod, further confirming that interannual variations in meteorological conditions had limited impact on the long-term EC trend. This also reinforces the reliability of the FEA method. Table S7 in the Supplement presents additional evidence: concentrations of co-emitted pollutants such as CO, SO₂, and NO₂ also declined significantly across all four cities from 2013–2023, corroborating that the downward trend in EC was predominantly driven by emission control efforts rather than meteorological variability. Diurnal patterns of EC, shown in Fig. S11a–d, consistently revealed peak concentrations during the morning and evening rush hour. However, the magnitude of these peaks has declined markedly in recent years. Correspondingly, Fig. S11e–h show the diurnal profile of emission control-driven EC relative to 2013, highlighting substantial reductions during rush-hour periods – a clear indication of the effectiveness of vehicle emission control policies.

Figure S6a–d and Table S7 further demonstrate strong correlations between EC and NO₂ concentrations, suggesting that vehicle emissions are a major contributor to urban BC levels. The observed reductions in NO₂ in all four cities underscore the impact of transportation-related emission control measures on declining EC concentrations. We further examined rush-hour EC trends in Figs. S14 and S15 in the Supplement and found more pronounced declines during these periods compared to the interannual average, indicating that targeted control strategies for mobile sources have been particularly effective. This finding aligns with previous work (e.g., Zheng et al., 2018), which reported substantial reductions in BC emissions from the transportation sector in China between 2010 and 2017.

Following the launch of China's Air Pollution Prevention and Control Action Plan in 2013, significant advances in air quality management were achieved. For example, Zhang et al. (2019a) attributed substantial decreases in PM_{2.5} concentrations between 2013 and 2017 mainly to reductions in anthropogenic emissions. To assess the impact of different policy phases (Geng et al., 2024; Zhang et al., 2019a), we divided the 11-year study period into three stages: 2013–2017, 2018–2020, and 2021–2023 (see Fig. 5). During the first stage (2013–2017), cities in the Yangtze River Delta implemented targeted control strategies. Nanjing and Suzhou, as economically developed industrial centers, prioritized the management of industrial parks and vehicle emission standards. Xuzhou, dominated by coal and heavy industries, emphasized curbing coal consumption and industrial emissions (Guo et al., 2022; Zhang et al., 2018). These coordinated efforts led to considerable reductions in EC: concentrations decreased by 36 % in Nanjing, 21 % in Suzhou, 18 % in Xuzhou, and 6 % in Zhenjiang. The smaller reduction observed in Zhenjiang may be attributed to its lower initial EC levels, limiting the absolute effect of control measures. Nonetheless, anthropogenic emissions remained the dominant driver of EC reductions, contributing 96 %, 99 %, 96 %, and 98 % of the observed changes in the four cities,

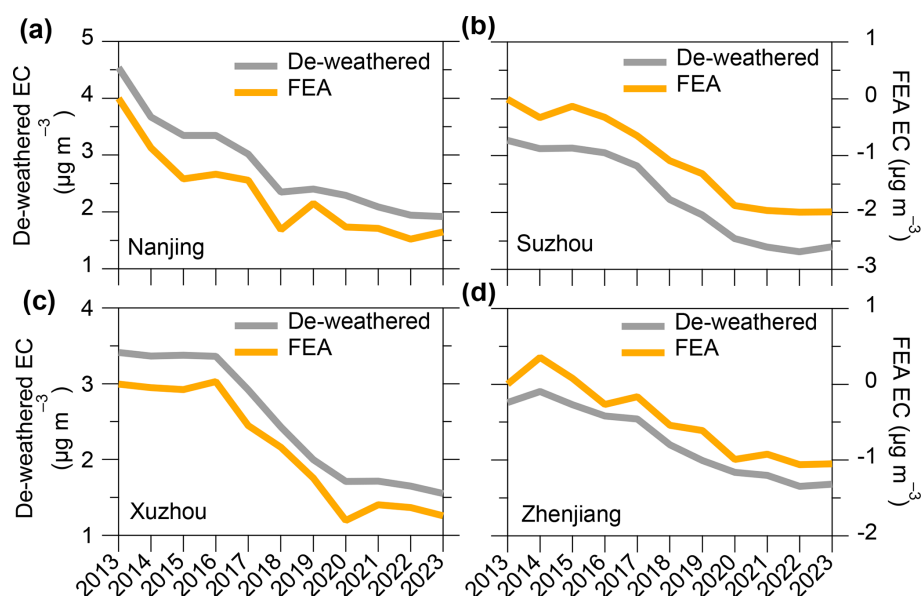


Figure 4. Trend in emission-driven EC from 2013–2023. Comparison of the results obtained from two different methods, including the FEA method developed in this study and the widely used de-weathered method.

respectively. During the second stage (2018–2020), further progress was achieved under the Three-Year Action Plan for Clean Air. EC concentrations declined by 24 % (Nanjing), 47 % (Suzhou), 44 % (Xuzhou), and 34 % (Zhenjiang). Average EC reductions attributed to anthropogenic emissions were $-0.82 \mu\text{g m}^{-3}$ in Nanjing and $-1.25 \mu\text{g m}^{-3}$ in Xuzhou. These reductions were likely amplified by the COVID-19 lockdowns, which significantly curtailed industrial activities and transportation in these cities. As shown in Fig. 5, we quantified the impact of the lockdown on EC levels, estimating that the pandemic contributed 3 %–10 % of the total EC reduction from 2017 to 2020 – an effect stemming from both emission reductions and changes in human activity. This is consistent with Zheng et al. (2021), who reported large-scale BC emission reductions during the lockdowns of early 2020, especially from industrial, residential, and transportation sources. In the third stage (2021–2023), the rate of EC decline began to slow. Reductions were 20 % in Nanjing, 7 % in Suzhou, 11 % in Xuzhou, and 14 % in Zhenjiang, primarily due to diminishing marginal returns from existing emission control measures. Meteorological influences during this period remained consistent with their contribution over the full 11-year span. Nevertheless, anthropogenic emissions still accounted for 93 % (Nanjing), 68 % (Suzhou), 67 % (Xuzhou), and 86 % (Zhenjiang) of the EC decline from 2020–2023.

4 Conclusion and implications

This study developed an ensemble learning approach to reconstruct hourly long-term in situ elemental carbon (EC) data using meteorological parameters and co-emitted source in-

dicators. Applied to four cities in the Yangtze River Delta (YRD) region of eastern China, the ensemble model demonstrated superior performance compared to individual algorithms – outperforming XGBoost, GBRT, and random forest by 17 %, 13 %, and 8 %, respectively. On average, the reconstructed dataset successfully filled 55 % of the missing EC observations. Its reliability was validated through consistency with ground-based EC, CO, and NO₂ measurements, as well as with TAP BC observations. The proposed approach provides an efficient and scalable solution to address missing EC data in long-term observational records. It also offers a means to reduce uncertainties in satellite-derived and reanalysis-based BC datasets. Over the 11-year study period, EC concentrations exhibited an overall declining trend, ranging from -0.20 to $-0.14 \mu\text{g m}^{-3} \text{a}^{-1}$. A more rapid decline was observed during 2013–2020 (-0.24 to $-0.15 \mu\text{g m}^{-3} \text{a}^{-1}$), followed by a notable deceleration between 2020 and 2023 (-0.12 to $-0.04 \mu\text{g m}^{-3} \text{a}^{-1}$).

To disentangle the drivers of EC trends, we developed the FEA method based on the ensemble model to quantify the respective contributions of anthropogenic emissions and meteorological variability. The results indicate that emission controls accounted for 91 % of the overall EC decline from 2013–2023, underscoring their dominant role. The slowed decline during 2020–2023 was mainly attributed to weakened contributions from anthropogenic emission reductions, which ranged from 67 % to 93 % across the four cities. The ensemble learning framework proposed in this study offers a robust and generalizable tool for reconstructing incomplete air quality datasets, with potential applications to a broad range of atmospheric pollutants. Furthermore, the integration of machine-learning-based reconstruction with the FEA

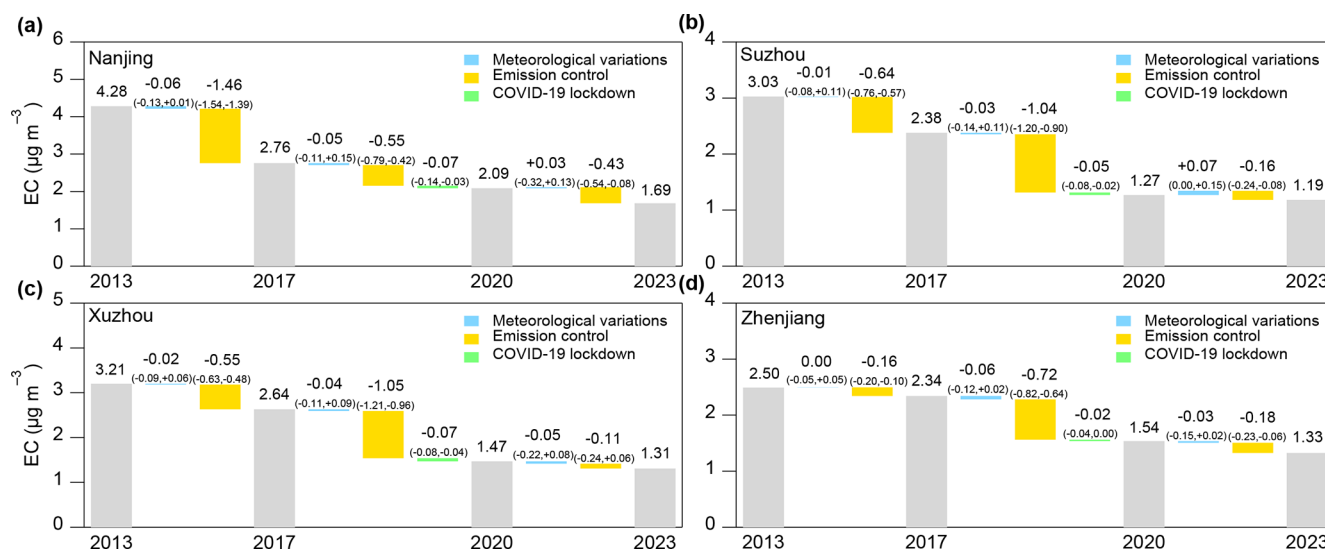


Figure 5. Drivers of the EC trend from 2013–2023. (a–d) The contributions of anthropogenic emission control, meteorological variations, and COVID-19 lockdown on the trends in EC concentration in the four cities (the values in parentheses represent the minimum to maximum values of the driving contributions analyzed by the FEA method).

method presented an effective approach for trend attribution – particularly for long-term observational records. While this method has proven effective for trend analysis of primary particulate matter, such as EC, in this study, its applicability to other atmospheric constituents with different sources and data characteristics remains to be evaluated. Future work could explore its performance and quantify associated uncertainties across diverse species and datasets.

Overall, our findings highlight the substantial impact of emission control policies on mitigating urban particulate EC pollution in the YRD. To sustain and accelerate reductions in EC and broader black carbon concentrations, continued enhancement of emission control measures, technological innovations, and policy enforcement will be essential. These efforts will be crucial not only for improving urban air quality but also for mitigating associated climate impacts.

Data availability. The MERRA-2 (M2T1NXADG, V5.12.4) data can be downloaded at <https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/M2T1NXAER.5.12.4/> (Gelaro et al., 2017).

The ERA5 reanalysis data can be downloaded at <https://doi.org/10.24381/cds.bd0915c6> (Hersbach et al., 2023).

The TAP data are from <http://tapdata.org.cn/> (Geng et al., 2021). The additional data will be made available upon request (yijzhang@nuist.edu.cn).

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/acp-25-7485-2025-supplement>.

Author contributions. YZ initiated and designed the study and developed the statistical methodology. YZ, QM, and JF developed the model code. QM and JF performed the simulations and analysis. YZ, SZ, YR, JQ, LT, and MZ conducted field measurements and validated the data. QM and YZ prepared the original manuscript, and XX, JEP, OF, and XG provided comments on the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (grant no. 42207124) and the Natural Science Foundation of Jiangsu Province (grant no. BK20210663).

Financial support. This research has been supported by the National Natural Science Foundation of China (grant no. 42207124) and the Natural Science Foundation of Jiangsu Province (grant no. BK20210663).

Review statement. This paper was edited by Dantong Liu and reviewed by four anonymous referees.

References

- Arhami, M., Kuhn, T., Fine, P. M., Delfino, R. J., and Sioutas, C.: Effects of sampling artifacts and operating parameters on the performance of a semicontinuous particulate elemental carbon/organic carbon monitor, *Environ. Sci. Technol.*, 40, 945–954, <https://doi.org/10.1021/es0510313>, 2006.
- Bali, K., Mishra, A. K., and Singh, S.: Impact of anomalous forest fire on aerosol radiative forcing and snow cover over Himalayan region, *Atmos. Environ.*, 150, 264–275, <https://doi.org/10.1016/j.atmosenv.2016.11.061>, 2017.
- Bao, F., Cheng, T., Li, Y., Gu, X., Guo, H., Wu, Y., Wang, Y., and Gao, J.: Retrieval of black carbon aerosol surface concentration using satellite remote sensing observations, *Remote Sens. Environ.*, 226, 93–108, <https://doi.org/10.1016/j.rse.2019.03.036>, 2019.
- Birch, M. and Cary, R.: Elemental carbon-based method for monitoring occupational exposures to particulate diesel exhaust, *Aerosol Sci. Tech.*, 25, 221–241, <https://doi.org/10.1080/02786829608965393>, 1996.
- Bond, T. C. and Bergstrom, R. W.: Light absorption by carbonaceous particles: An investigative review, *Aerosol Sci. Tech.*, 40, 27–67, <https://doi.org/10.1080/02786820500421521>, 2006.
- Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., and Koch, D.: Bounding the role of black carbon in the climate system: A scientific assessment, *J. Geophys. Res.-Atmos.*, 118, 5380–5552, <https://doi.org/10.1002/jgrd.50171>, 2013.
- Carslaw, D. C.: Evidence of an increasing NO₂/NO_x emissions ratio from road traffic emissions, *Atmos. Environ.*, 39, 4793–4802, 2005.
- Cross, E. S., Onasch, T. B., Ahern, A., Wrobel, W., Slowik, J. G., Olfert, J., Lack, D. A., Massoli, P., Cappa, C. D., and Schwarz, J. P.: Soot particle studies – instrument inter-comparison – project overview, *Aerosol Sci. Tech.*, 44, 592–611, 2010.
- Cui, S., Xian, J., Shen, F., Zhang, L., Deng, B., Zhang, Y., and Ge, X.: One-year real-time measurement of black carbon in the rural area of Qingdao, Northeastern China: Seasonal variations, meteorological effects, and the COVID-19 case analysis, *Atmosphere*, 12, 394, <https://doi.org/10.3390/atmos12030394>, 2021.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., and Bauer, D. P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- Ding, A., Huang, X., Nie, W., Sun, J., Kerminen, V. M., Petäjä, T., Su, H., Cheng, Y., Yang, X. Q., and Wang, M.: Enhanced haze pollution by black carbon in megacities in China, *Geophys. Res. Lett.*, 43, 2873–2879, <https://doi.org/10.1002/2016GL067745>, 2016.
- Ervens, B.: Modeling the processing of aerosol and trace gases in clouds and fogs, *Chem. Rev.*, 115, 4157–4198, <https://doi.org/10.1021/cr5005887>, 2015.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., and Reichle, R.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), *J. Climate*, 30, 5419–5454, 2017 (data available at: <https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/M2T1NXAER.5.12.4/>, last access: 20 February 2024).
- Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., Peng, Y., Huang, X., He, K., and Zhang, Q.: Tracking Air Pollution in China: Near Real-Time PM_{2.5} Retrievals from Multisource Data Fusion, *Environ. Sci. Technol.*, 55, 12106–12115, <https://doi.org/10.1021/acs.est.1c01863>, 2021 (data available at: <http://tapdata.org/cn/>, last access: 20 February 2024).
- Geng, G., Liu, Y., Liu, Y., Liu, S., Cheng, J., Yan, L., Wu, N., Hu, H., Tong, D., and Zheng, B.: Efficacy of China's clean air actions to tackle PM_{2.5} pollution between 2013 and 2020, *Nat. Geosci.*, 17, 987–994, 2024.
- Gogoi, M. M., Babu, S. S., Imasu, R., and Hashimoto, M.: Satellite (GOSAT-2 CAI-2) retrieval and surface (ARFINET) observations of aerosol black carbon over India, *Atmos. Chem. Phys.*, 23, 8059–8079, <https://doi.org/10.5194/acp-23-8059-2023>, 2023.
- Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM₁₀ trend analysis, *Atmos. Chem. Phys.*, 18, 6223–6239, <https://doi.org/10.5194/acp-18-6223-2018>, 2018.
- Guo, Y., Li, K., Zhao, B., Shen, J., Bloss, W. J., Azzi, M., and Zhang, Y.: Evaluating the real changes of air quality due to clean air actions using a machine learning technique: Results from 12 Chinese mega-cities during 2013–2020, *Chemosphere*, 300, 134608, <https://doi.org/10.1016/j.chemosphere.2022.134608>, 2022.
- Harrison, R. M.: Urban atmospheric chemistry: a very special case for study, *npj Climate and Atmospheric Science*, 1, 20175, <https://doi.org/10.1038/s41612-017-0010-8>, 2018.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], <https://doi.org/10.24381/cds.bd0915c6>, 2023.
- Hoerl, A. E. and Kennard, R. W.: Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, 12, 55–67, 1970.
- Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., and Nie, W.: Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China, *Natl. Sci. Rev.*, 8, nwaa137, <https://doi.org/10.1093/nsr/nwaa137>, 2021.
- Jung, J., Kim, Y. J., Lee, K. Y., Kawamura, K., Hu, M., and Kondo, Y.: The effects of accumulated refractory particles and the peak inert mode temperature on semi-continuous organic carbon and elemental carbon measurements during the CAREBeijing 2006 campaign, *Atmos. Environ.*, 45, 7192–7200, <https://doi.org/10.1016/j.atmosenv.2011.09.003>, 2011.
- Kidwell, J. S. and Brown, L. H.: Ridge regression as a technique for analyzing models with multicollinearity, *J. Marriage Fam.*, 287–299, 1982.
- Laj, P., Lund Myhre, C., Riffault, V., Amiridis, V., Fuchs, H., Eleftheriadis, K., Petäjä, T., Salameh, T., Kivekäs, N., and Jurola, E.: Aerosol, Clouds and Trace Gases Research Infrastructure – ACTRIS, the European research infrastructure supporting

- atmospheric science, *B. Am. Meteorol. Soc.*, 105, E1098–E1136, <https://doi.org/10.1175/BAMS-D-23-0064.1>, 2024.
- Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, *Atmos. Meas. Tech.*, 6, 2989–3034, <https://doi.org/10.5194/amt-6-2989-2013>, 2013.
- Li, L., Derimian, Y., Chen, C., Zhang, X., Che, H., Schuster, G. L., Fuertes, D., Litvinov, P., Lapyonok, T., Lopatin, A., Matar, C., Ducos, F., Karol, Y., Torres, B., Gui, K., Zheng, Y., Liang, Y., Lei, Y., Zhu, J., Zhang, L., Zhong, J., Zhang, X., and Dubovik, O.: Climatology of aerosol component concentrations derived from multi-angular polarimetric POLDER-3 observations using GRASP algorithm, *Earth Syst. Sci. Data*, 14, 3439–3469, <https://doi.org/10.5194/essd-14-3439-2022>, 2022.
- Li, Y., Lei, L., Sun, J., Gao, Y., Wang, P., Wang, S., Zhang, Z., Du, A., Li, Z., and Wang, Z.: Significant reductions in secondary aerosols after the three-year action plan in Beijing summer, *Environ. Sci. Technol.*, 57, 15945–15955, <https://doi.org/10.1021/acs.est.3c02417>, 2023.
- Liu, D., Flynn, M., Gysel, M., Targino, A., Crawford, I., Bower, K., Choularton, T., Jurányi, Z., Steinbacher, M., Hüglin, C., Curtius, J., Kampus, M., Petzold, A., Weingartner, E., Baltensperger, U., and Coe, H.: Single particle characterization of black carbon aerosols at a tropospheric alpine site in Switzerland, *Atmos. Chem. Phys.*, 10, 7389–7407, <https://doi.org/10.5194/acp-10-7389-2010>, 2010.
- Liu, S., Aiken, A. C., Gorkowski, K., Dubey, M. K., Cappa, C. D., Williams, L. R., Herndon, S. C., Massoli, P., Fortner, E. C., Chhabra, P. S., Brooks, W. A., Onasch, T. B., Jayne, J. T., Worsnop, D. R., China, S., Sharma, N., Mazzoleni, C., Xu, L., Ng, N. L., Liu, D., Allan, J. D., Lee, J. D., Fleming, Z. L., Mohr, C., Zotter, P., Szidat, S., and Prevot, A. S. H.: Enhanced light absorption by mixed source black and brown carbon particles in UK winter, *Nat. Commun.*, 6, 8435, <https://doi.org/10.1038/ncomms9435>, 2015.
- Liu, S., Geng, G., Xiao, Q., Zheng, Y., Liu, X., Cheng, J., and Zhang, Q.: Tracking Daily Concentrations of PM_{2.5} Chemical Composition in China since 2000, *Environ. Sci. Technol.*, 56, 16517–16527, <https://doi.org/10.1021/acs.est.2c06510>, 2022.
- Liu, Z., Deng, Z., Davis, S. J., and Ciais, P.: Global carbon emissions in 2023, *Nature Reviews Earth and Environment*, 1–2, <https://doi.org/10.1038/s43017-024-00532-2>, 2024.
- Ma, X., Yan, P., Zhao, T., Jia, X., Jiao, J., Ma, Q., Wu, D., Shu, Z., Sun, X., and Habtemicheal, B. A.: Evaluations of Surface PM₁₀ concentration and chemical compositions in MERRA-2 aerosol reanalysis over Central and Eastern China, *Remote Sens.*, 13, 1317, <https://doi.org/10.3390/rs13071317>, 2021.
- Matsui, H.: Black carbon absorption efficiency under preindustrial and present-day conditions simulated by a size- and mixing-state-resolved global aerosol model, *J. Geophys. Res.-Atmos.*, 125, e2019JD032316, <https://doi.org/10.1029/2019JD032316>, 2020.
- Ng, N. L., Dillner, A. M., Bahreini, R., Russell, A. G., de La Beaujardière, J., Flynn, J., Gentner, D. R., Griffin, R., Hawkins, L. N., and Jimenez, J. L.: Atmospheric Science and Chemistry Measurement NeTwork (ASCENT): A new ground-based high time-resolution air quality monitoring network, *AGU Fall Meeting Abstracts*, 1 December 2022, Chicago, IL, A55R-1360, <https://ui.adsabs.harvard.edu/abs/2022AGUFM.A55R1360N> (last access: 20 February 2024), 2022.
- Peng, X., Liu, M., Zhang, Y., Meng, Z., Achal, V., Zhou, T., Long, L., and She, Q.: The characteristics and local-regional contributions of atmospheric black carbon over urban and suburban locations in Shanghai, China, *Environ. Pollut.*, 255, 113188, <https://doi.org/10.1016/j.envpol.2019.113188>, 2019.
- Pileci, R. E., Modini, R. L., Bertò, M., Yuan, J., Corbin, J. C., Marinoni, A., Henzing, B., Moerman, M. M., Putaud, J. P., Spindler, G., Wehner, B., Müller, T., Tuch, T., Trentini, A., Zanatta, M., Baltensperger, U., and Gysel-Beer, M.: Comparison of co-located refractory black carbon (rBC) and elemental carbon (EC) mass concentration measurements during field campaigns at several European sites, *Atmos. Meas. Tech.*, 14, 1379–1403, <https://doi.org/10.5194/amt-14-1379-2021>, 2021.
- Ramanathan, V. and Carmichael, G.: Global and regional climate changes due to black carbon, *Nat. Geosci.*, 36, 335–358, <https://doi.org/10.1038/ngeo156>, 2008.
- Randles, C., Da Silva, A., Buchard, V., Colarco, P., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., and Hair, J.: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation, *J. Climate*, 30, 6823–6850, <https://doi.org/10.1175/JCLI-D-16-0609.1>, 2017.
- Reid, J. S., Koppmann, R., Eck, T. F., and Eleuterio, D. P.: A review of biomass burning emissions part II: intensive physical properties of biomass burning particles, *Atmos. Chem. Phys.*, 5, 799–825, <https://doi.org/10.5194/acp-5-799-2005>, 2005.
- Valavanidis, A., Vlachogianni, T., Fiotakis, K., and Lioridas, S.: Pulmonary oxidative stress, inflammation and cancer: respirable particulate matter, fibrous dusts and ozone as major causes of lung carcinogenesis through reactive oxygen species mechanisms, *Int. J. Env. Res. Pub. He.*, 10, 3886–3907, <https://doi.org/10.3390/ijerph10093886>, 2013.
- Vu, T. V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., and Harrison, R. M.: Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique, *Atmos. Chem. Phys.*, 19, 11303–11314, <https://doi.org/10.5194/acp-19-11303-2019>, 2019.
- Wang, Q., Jacob, D. J., Fisher, J. A., Mao, J., Leibensperger, E. M., Carouge, C. C., Le Sager, P., Kondo, Y., Jimenez, J. L., Cubison, M. J., and Doherty, S. J.: Sources of carbonaceous aerosols and deposited black carbon in the Arctic in winter-spring: implications for radiative forcing, *Atmos. Chem. Phys.*, 11, 12453–12473, <https://doi.org/10.5194/acp-11-12453-2011>, 2011.
- Wang, S. and Chen, B.: Accounting of SO₂ emissions from combustion in industrial boilers, *Energy Procedia*, 88, 325–329, 2016.
- Wei, C., Wang, M., Fu, Q., Dai, C., Huang, R., and Bao, Q.: Temporal characteristics and potential sources of black carbon in megacity Shanghai, China, *J. Geophys. Res.-Atmos.*, 125, e2019JD031827, <https://doi.org/10.1029/2019JD031827>, 2020.
- Wei, J., Wang, J., Li, Z., Kondragunta, S., Anenberg, S., Wang, Y., Zhang, H., Diner, D., Hand, J., and Lyapustin, A.: Long-term mortality burden trends attributed to black carbon and PM_{2.5} from wildfire emissions across the continental USA from 2000 to 2020: a deep learning modelling study, *Lancet Planetary Health*, 7, e963–e975, [https://doi.org/10.1016/S2542-5196\(23\)00235-8](https://doi.org/10.1016/S2542-5196(23)00235-8), 2023.

- Xu, H., Ren, Y. a., Zhang, W., Meng, W., Yun, X., Yu, X., Li, J., Zhang, Y., Shen, G., and Ma, J.: Updated global black carbon emissions from 1960 to 2017: improvements, trends, and drivers, *Environ. Sci. Technol.*, 55, 7869–7879, <https://doi.org/10.1021/acs.est.1c03117>, 2021.
- Xu, X., Yang, X., Zhu, B., Tang, Z., Wu, H., and Xie, L.: Characteristics of MERRA-2 black carbon variation in east China during 2000–2016, *Atmos. Environ.*, 222, 117140, <https://doi.org/10.1016/j.atmosenv.2019.117140>, 2020.
- Yang, J., Ji, Z., Kang, S., and Tripathee, L.: Contribution of South Asian biomass burning to black carbon over the Tibetan Plateau and its climatic impact, *Environ. Pollut.*, 270, 116195, <https://doi.org/10.1016/j.envpol.2020.116195>, 2021.
- Yang, Y., Xu, X., Zhang, Y., Zheng, S., Wang, L., Liu, D., Gustave, W., Jiang, L., Hua, Y., and Du, S.: Seasonal size distribution and mixing state of black carbon aerosols in a polluted urban environment of the Yangtze River Delta region, China, *Sci. Total Environ.*, 654, 300–310, 2019.
- Yu, H., Li, M., Zheng, X., Zhu, M., Zheng, Z., Xie, T., Yan, G., Hu, P., Cao, Z., and Feng, J.: Potential source and health risks of black carbon based on MERRA-2 reanalysis data in a typical industrial city of North China Plain, *J. Environ. Manage.*, 354, 120367, <https://doi.org/10.1016/j.jenvman.2024.120367>, 2024.
- Zhang, F., Cheng, H.-R., Wang, Z.-W., Lv, X.-P., Zhu, Z.-M., Zhang, G., and Wang, X.-M.: Fine particles (PM_{2.5}) at a CAWNET background site in Central China: Chemical compositions, seasonal variations and regional pollution events, *Atmos. Environ.*, 86, 193–202, <https://doi.org/10.1016/j.atmosenv.2013.12.008>, 2014.
- Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., and Liu, W.: Drivers of improved PM_{2.5} air quality in China from 2013 to 2017, *P. Natl. Acad. Sci. USA*, 116, 24463–24469, 2019a.
- Zhang, S., Ren, H., Zhou, W., Yu, Y., and Chen, C.: Assessing air pollution abatement co-benefits of energy efficiency improvement in cement industry: A city level analysis, *J. Clean. Prod.*, 185, 761–771, 2018.
- Zhang, Y., Vu, T. V., Sun, J., He, J., Shen, X., Lin, W., Zhang, X., Zhong, J., Gao, W., and Wang, Y.: Significant changes in chemistry of fine particles in wintertime Beijing from 2007 to 2017: impact of clean air actions, *Environ. Sci. Technol.*, 54, 1344–1352, <https://doi.org/10.1021/acs.est.9b04678>, 2019b.
- Zhao, J., Liu, Y., Shan, M., Liang, S., Cui, C., Chen, L., Gao, S., Mao, J., Zhang, H., and Sun, Y.: Characteristics, potential regional sources and health risk of black carbon based on ground observation and MERRA-2 reanalysis data in a coastal city, China, *Atmos. Res.*, 256, 105563, <https://doi.org/10.1016/j.atmosres.2021.105563>, 2021.
- Zhao, J., Chen, H., Qi, X., Chi, X., Jia, M., Jiang, F., Zhong, S., Zheng, B., and Ding, A.: Observed decade-long improvement of combustion efficiency in the Yangtze River Delta region in China, *Environ. Res. Lett.*, 19, 074001, <https://doi.org/10.1088/1748-9326/ad521e>, 2024.
- Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, *Atmos. Chem. Phys.*, 18, 14095–14111, <https://doi.org/10.5194/acp-18-14095-2018>, 2018.
- Zheng, B., Zhang, Q., Geng, G., Chen, C., Shi, Q., Cui, M., Lei, Y., and He, K.: Changes in China's anthropogenic emissions and air quality during the COVID-19 pandemic in 2020, *Earth Syst. Sci. Data*, 13, 2895–2907, <https://doi.org/10.5194/essd-13-2895-2021>, 2021.
- Zheng, G. J., Cheng, Y., He, K. B., Duan, F. K., and Ma, Y. L.: A newly identified calculation discrepancy of the Sunset semi-continuous carbon analyzer, *Atmos. Meas. Tech.*, 7, 1969–1977, <https://doi.org/10.5194/amt-7-1969-2014>, 2014.
- Zhou, W., Lei, L., Du, A., Zhang, Z., Li, Y., Yang, Y., Tang, G., Chen, C., Xu, W., and Sun, J.: Unexpected increases of severe haze pollution during the post COVID-19 period: Effects of emissions, meteorology, and secondary production, *J. Geophys. Res.-Atmos.*, 127, e2021JD035710, <https://doi.org/10.1029/2021JD035710>, 2022.
- Zhou, Y., Zhuang, B., Wang, T., Gao, P., Li, S., Hu, Y., Li, M., Cao, H., Xie, M., and Chen, H.: Characteristics of urban black carbon aerosols in the Yangtze River Delta of China based on long-term observations, *Atmos. Environ.*, 326, 120488, <https://doi.org/10.1016/j.atmosenv.2024.120488>, 2024.